On the distribution of depths in increasing trees.

Markus Kuba
Institut für Diskrete Mathematik und Geometrie
Technische Universität Wien
Wiedner Hauptstr. 8-10/104, 1040 Wien, Austria
kuba@dmg.tuwien.ac.at

and

Stephan Wagner
Department of Mathematical Sciences
Stellenbosch University
Private Bag X1, Matieland 7602, South Africa
swagner@sun.ac.za

Mathematics Subject Classifications: 05A19; 05C05 60C05

Abstract

By a theorem of Dobrow and Smythe, the depth of the kth node in very simple families of increasing trees (which includes, among others, binary increasing trees, recursive trees and plane ordered recursive trees) follows the same distribution as the number of edges of the form $j-(j+1)$ with $j<k$. In this short note, we present a simple bijective proof of this fact, which also shows that the result actually holds within a wider class of increasing trees. We also discuss some related results that follow from the bijection as well as a possible generalization.

1 Introduction

Increasing trees are rooted labeled trees where the nodes of a tree of size n are labeled by distinct integers from the set $\{1,\ldots,n\}$ in such a way that the sequence of labels along any branch starting at the root is increasing. There are various important families of increasing trees, such as *binary increasing trees*, *recursive trees* or *plane-oriented recursive trees*. A general framework for these instances is given by what is known as *simple families of increasing trees* [3]; such a family $\mathcal{T}$ is characterized by a sequence of non-negative numbers $(\varphi_k)_{k\geq 0}$, where $\varphi_0 > 0$. This sequence is called the degree-weight sequence. We assume that there exists a $k \geq 2$ with $\varphi_k > 0$ to avoid trivialities.

Now we assign a *weight* $w(T)$ to any ordered tree T by $w(T) := \prod_v \varphi_{d(v)}$, where v ranges over all vertices of T and $d(v)$ is the out-degree of v. Furthermore, let $L(T)$ be the number of increasing labelings of T with integers $1, 2, \ldots, |T|$, as explained above, and define the total weights by $T_n := \sum_{|T|=n} w(T) \cdot L(T)$. It follows that the exponential generating function $T(z) := \sum_{n\geq 1} T_n \frac{z^n}{n!}$ satisfies the autonomous first order differential equation

$$T'(z) = \Phi\big(T(z)\big), \quad T(0) = 0, \tag{1}$$

where $\Phi(t) = \sum_{n=0}^{\infty} \varphi_n t^n$. This equation follows easily from the fact that one can describe a tree as a root node with several subtrees from the same family attached to it (see for instance [3] or [4]).

Important special cases include $\Phi(t) = 1+t^2$, which corresponds to binary increasing trees, $\Phi(t) = e^t$ (recursive trees), and $\Phi(t) = \frac{1}{1-t}$ (plane-oriented recursive trees. In all these cases, the total weight can simply be interpreted as the number of trees of given size within the family. Binary trees are essentially equivalent to binary search trees, which in turn serve as an analytic model for the famous Quicksort algorithm [7]. Plane-oriented recursive trees, on the other hand, are a special instance of the well known Barabási-Albert model [2] for scale-free networks (see also [5]), which is used as a simplified growth model of the world wide web [1].

From a combinatorial point of view, it is interesting to note that binary increasing trees are enumerated by the tangent numbers (see [9] for various interesting bijections), while there are $(n-1)!$ recursive trees and $(2n-3)!!$ plane-oriented recursive trees with n nodes.

A specific subclass of increasing trees is known as *very simple families* [10] of increasing trees. The three aforementioned examples all belong to this subclass, which is essentially characterized by the fact that the function $\Phi(t)$ is either of the form $(1 + ct)^\alpha$ for constants c, α of the same sign or of the form e^{ct} for some positive constant c. These specific families have the property that they can be described via a tree evolution process, as pointed out by Panholzer and Prodinger in [10].

A remarkable result by Dobrow and Smythe [6] states that the depth of the kth node (i.e., the distance from the root) in a random increasing tree from one of the very simple families follows the exact same distribution as the number of edges between two vertices whose labels are $\leq k$ and differ by exactly 1 (henceforth, we will simply call such edges 1-*edges*). See also [10]. The aim of this short note is to show that this holds more generally for simple families of increasing trees, and to present a simple bijective proof of this fact. Several further corollaries follow as well, and the bijection can also be generalized, see Section 3.

2 The bijection

Let us now describe a bijection B_k on the set of ordered increasing trees as follows:

- If node $j-1$ lies on the unique path from 1 to k in T and ℓ is its successor on this path, then j takes the position of ℓ in $B_k(T)$ (i.e., it is attached to $j-1$ in the same position as ℓ in T).

- If $j \leq k$ but node $j-1$ does not lie on this path, then j takes the position of $j-1$ in $B_k(T)$.

- If $j > k$, then the positions of j in T and $B_k(T)$ are the same.

The inverse operation B_k^{-1} is equally simple:

- If vertices j and $j+1$ are connected in T, then j lies on the path from 1 to k in $B_k^{-1}(T)$, and the successor of j on this path takes the position of $j+1$.

- If $j < k$ but vertices j and $j+1$ are not connected, then the position of j in $B_k^{-1}(T)$ is the same as the position of $j+1$ in T.

- If $j > k$, then the positions of j in T and $B_k^{-1}(T)$ are the same.

It is easy to see that both operations are well-defined and inverses of each other. Figure 1 shows an example with $k = 9$.

The following properties of the bijection are immediate:

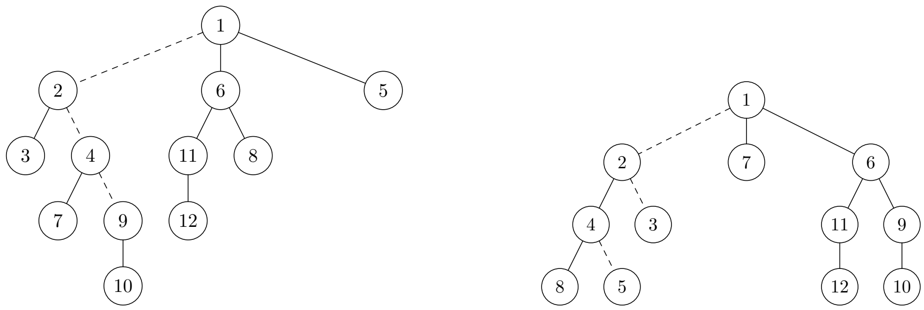

Figure 1: The bijection in an example: T (left) and $B_9(T)$ (right).

- For any increasing tree T with $n \geq k$ nodes, $B_k(T)$ is an increasing tree with n nodes and the same outdegrees.
- Edges on the path between the root and k are mapped to 1-edges in $B_k(T)$ whose ends are labeled with numbers $\leq k$.

Since all outdegrees remain the same, the weights $w(T)$ and $w(B_k(T))$ are also always the same, regardless of the degree-weight sequence φ. The following results are obtained as a consequence. For very simple families of increasing trees, these theorems occur in the aforementioned paper by Dobrow and Smythe [6]:

Theorem 1 (Dobrow/Smythe, Theorem 5)
In a random increasing tree with n nodes from a simple family, the probability that k is attached to j is exactly the probability that the last 1*-edge with labels* $\leq k$ *is between* j *and* $j+1$.

More generally, the following holds:

Theorem 2 *In a random increasing tree with n nodes from a simple family, the probability that the ancestors of k are* $j_1, j_2, \ldots, j_s$ *in this order* ($j_1 > j_2 > \cdots > j_s$) *is the same as the probability that the only* 1*-edges with labels between* j_s *and* k *are* $j_1 - (j_1+1)$, $j_2 - (j_2+1)$, $\ldots$, $j_s - (j_s+1)$.

Theorem 3 (Dobrow/Smythe, Theorem 7)
In a random increasing tree with n nodes from a simple family, the distribution of the depth of node k is the same as the distribution of the number of 1*-edges with labels* $\leq k$. *Furthermore, the probability that node j lies on the*

unique path between 1 *and* k *is the same as the probability that there is an edge between* j *and* $j+1$*.*

In particular, one has the following corollary:

Corollary 4 *The probability that* j *lies on the path between* 1 *and* k *does not depend on* k*.*

Remark 1 None of the above theorems depends on the size of the increasing tree. In the case of very simple families, which can be generated by a growth process, this is essentially trivial, but it is quite surprising that this remains true within the more general setting of simple families of increasing trees.

3 Generalization and conclusion

Our bijection provides a simple combinatorial explanation for several results that were obtained in [6] by probabilistic techniques, with the additional benefit that they generalize to a wider range of increasing trees, namely to all simple families. The bijection can be generalized further to prove the following:

Theorem 5 (Dobrow/Smythe, Theorem 6)
In a random increasing tree with n *nodes from a simple family, the distribution of the distance between nodes* i *and* k $(i < k)$ *is the same as the distribution of the sum of the distance between* i *and* $i+1$ *and the number of number of* 1*-edges with labels between* $i+1$ *and* k*.*

To this end, consider a bijection $B_{i,k}$ that is defined as follows:

- If $i+1 < j$, node $j-1$ lies on the unique path from 1 to k in T and ℓ is its successor on this path, then j takes the position of ℓ in $B_{i,k}(T)$.
- If $i+1 < j \leq k$ but node $j-1$ does not lie on this path, then j takes the position of $j-1$ in $B_{i,k}(T)$.
- Vertex $i+1$ takes the position of the node in T that lies on the path between i and k and has the smallest label $> i$.
- If $j \leq i$ or $j > k$, then the positions of j in T and $B_{i,k}(T)$ are the same.

See Figure 2 for an example with $i = 4$ and $k = 12$. Note that the path between i and k is mapped to the path between i and $i + 1$ and a collection of 1-edges, thereby proving Theorem 5.

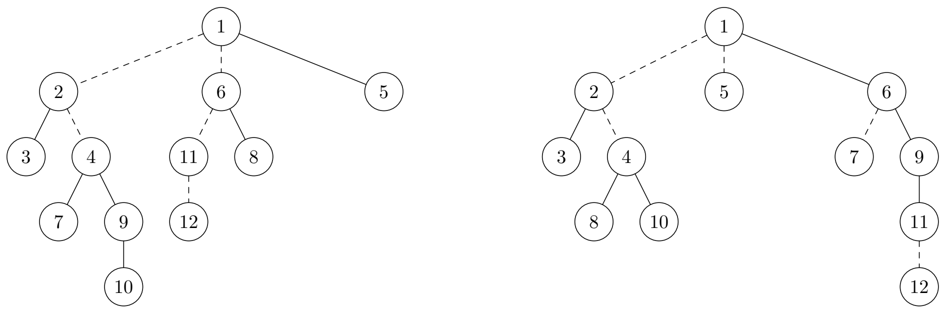

Figure 2: The generalized bijection in an example: T (left) and $B_{4,12}(T)$ (right).

Unfortunately it seems that, even though we know now that the distribution of the depths and distances of nodes is related to the number of 1-edges, it remains difficult to obtain precise results on this distribution if the variety under consideration is none of the very simple families, cf. [8, 10]. However, we believe that the bijection presented in this note, and similar ones, might also lead to new distributional results on distance-related parameters in increasing trees.

References

[1] R. Albert, H. Jeong, and A.-L. Barabási. The diameter of the world wide web. *Nature*, 401:130–131, 1999.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] F. Bergeron, P. Flajolet, and B. Salvy. Varieties of increasing trees. In *CAAP '92 (Rennes, 1992)*, volume 581 of *Lecture Notes in Comput. Sci.*, pages 24–48. Springer, Berlin, 1992.

[4] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial species and tree-like structures.* Cambridge University Press, 1998., Cambridge, 1998.

[5] B. Bollobás and O. M. Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34. Wiley-VCH, Weinheim, 2003.

[6] R. P. Dobrow and R. T. Smythe. Poisson approximations for functionals of random trees. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 79–92, 1996.

[7] C. A. R. Hoare. Quicksort. *Comput. J.*, 5:10–15, 1962.

[8] M. Kuba and A. Panholzer. On the distribution of distances between specified nodes in increasing trees. *Submitted.*

[9] A. G. Kuznetsov, I. M. Pak, and A. E. Postnikov. Increasing trees and alternating permutations. *Uspekhi Mat. Nauk*, 49(6(300)):79–110, 1994.

[10] A. Panholzer and H. Prodinger. Level of nodes in increasing trees revisited. *Random Structures Algorithms*, 31(2):203–226, 2007.